AD-A041 028 MISSOURI UNIV-COLUMBIA DEPT OF STATISTICS F/G 14/4

FAILURE RATE ESTIMATION USING RANDOM SMOOTHING.(U)

APR 77 D R MILLER, N D SINGPURWALLA N00014-76-C-0843

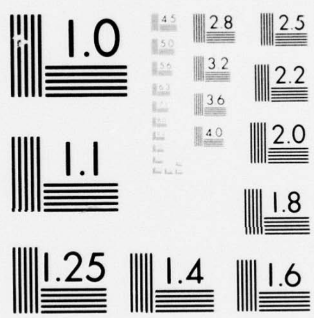
UNCLASSIFIED TR-67 NL

1 OF 1

AD
A041 028

END

DATE
FILMED
7-77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

University of Missouri-Columbia

See 1473 in back

Failure Rate Estimation Using Random Smoothing

by
Douglas R. Miller, University of Missouri
and
Nozer D. Singpurwalla, The George Washington University

Technical Report No. 67
Department of Statistics
April 1977

Mathematical

Sciences

FAILURE RATE ESTIMATION USING RANDOM SMOOTHING

by

Douglas R. Miller[1]
Statistics Department
University of Missouri
Columbia, Missouri 65201

and

Nozer D. Singpurwalla[2]
The George Washington University
Washington, D. C. 20052

April 1977

1. Research Sponsored by Contract N00014-76-C-0843, Office of Naval Research.

2. Research Jointly Sponsored by Contract AT(49-24)-0211, Nuclear Regulatory Commission, and Contract N00014-75-C-0729, Project NR 347-020, Office of Naval Research.

FAILURE RATE ESTIMATION USING RANDOM SMOOTHING

by

Douglas R. Miller
Nozer D. Singpurwalla

In this note, we investigate some aspects of the distribution theory of an estimator of the failure-rate function. We clarify and give a slightly different approach to some asymptotic results of Singpurwalla (1975) for the failure rate process from exponential lifetimes. In addition, we extend these results to the case of general lifetimes. This leads us to confidence regions for "randomly smoothed" versions of the failure-rate function, for both finite and large samples.

Research Jointly Sponsored by
Nuclear Regulatory Commission
and
Office of Naval Research

THE GEORGE WASHINGTON UNIVERSITY
School of Engineering and Applied Science
Institute for Management Science and Engineering

FAILURE RATE ESTIMATION USING RANDOM SMOOTHING

by

Douglas R. Miller
Nozer D. Singpurwalla

1. Introduction

Let F be the distribution function of a random lifetime, and let $r = -\frac{d}{dt}\log(1-F)$ be its failure rate function. The purpose of this note is to consider estimators of r based on complete and censored data, and to investigate the related distribution theory for both the finite and the infinite sample cases.

Consider a collection of n identical items which function independently of one another. Let X_i (L_i) be the time to failure (withdrawal) of the ith item. We assume that the X_i's are independent and identically distributed with distribution function F. The L_i's are random withdrawal times which may have any joint distribution, but must be independent of the X_i's. Let $Z_i = \min(X_i, L_i)$; thus, if $Z_i = X_i$, then the ith item has failed at X_i, and if $Z_i = L_i$, then the ith item was withdrawn from observation at L_i. Suppose that k failures have been observed in all, and let

$$0 \equiv Z_{(0)} \leq Z_{(1)} \leq Z_{(2)} \cdots \leq Z_{(k)}$$

be the ordered failure times. Let $N_n(t)$ be the total number of items on test at time t, and let $T_n(t) = \int_0^t N_n(u)du$ be the "total time on test" at time t.

We shall first define a "naive" estimator of r, say R_n, as

$$R_n(z) = \frac{1}{T_n(Z_{(i)}) - T_n(Z_{(i-1)})}, \quad Z_{(i-1)} < z \leq Z_{(i)} \qquad (1.1)$$
$$= \infty \qquad\qquad, \quad z > Z_{(k)} .$$

Some Comments on the Naive Estimator

We note that the estimator R_n is the reciprocal of the total time on test, on an interval over which it is defined. The estimator, in effect, assumes a piece-wise exponential distribution between successive failure points. In order to compute $R_n(z)$, we must look "ahead" to the failure which immediately follows z. Thus, the estimator $R_n(z)$ is a left-continuous step function.

Assuming that there is no censoring or withdrawals (i.e., when $k=n$) and assuming that n is large, Sethuraman and Singpurwalla (1977) show that $R_n(z)$ is not consistent. They induce consistency by smoothing this estimator using windows of fixed width. The width of their windows depends, among other things, on the sample size n. In this paper, we shall consider another smoothing procedure. However, we first obtain a right-continuous, closed form version of our estimator (1.1).

Right-Continuous Version of the Naive Estimator

Let F_n denote the empirical distribution function of the $Z_{(i)}$'s, and let

$$F_n^{-1}(u) = \inf\{x : F(x) > u\}$$

be the right-continuous inverse of F_n , with $F_n^{-1}(0) \equiv 0$. Then, for $Z_{(i-1)} \leq z < Z_{(i)}$,

$$F_n^{-1}F_n(z) = Z_{(i)} \quad \text{and} \quad F_n^{-1}(F_n(z) - k^{-1}) = Z_{(i-1)} .$$

Thus, given k failures, a right-continuous version of our naive estimator, in closed form, is

$$R_n(z) = \frac{1}{T_n(F_n^{-1}F_n(z)) - T_n(F_n^{-1}(F_n(z) - k^{-1}))} . \tag{1.2}$$

An analysis of the right- and the left-continuous versions of $R_n(z)$ will be the same, and thus we will not make any distinction of this throughout our paper. However, the notation in Sections 3 and 4 will be simplified by using right-continuity.

1.1 A Smoothed Version of the Naive Estimator

We next introduce what we call a "randomly smoothed" version of our naive estimator. Our motivation for choosing this smoothing technique is given in Section 1.2.

For $0 < h < 1$, let

$$[hk(n)] \overset{\text{def}}{=} \min\{i, \text{ an integer} : i \geq hk(n)\} ;$$

that is, $[hk(n)]$ is the smallest integer greater than or equal to $hk(n)$ where $k(n)$, the number of observed failures, is random. For $Z_{(i-1)} \leq z < Z_{(i)}$, we define a smoothed estimator $R_{n,h}(z)$ as

$$R_{n,h}(z) = \frac{[hk(n)]}{T_n(Z_{(i)}) - T_n(Z_{(i-[hk(n)])})} .$$

In closed form, the above estimator can also be written as

$$R_{n,h}(z) = \frac{[hk(n)]}{T_n(F_n^{-1}F_n(z)) - T_n(F_n^{-1}(F_n(z)-h))} . \tag{1.3}$$

If we set $j = [hk(n)]$, then for $Z_{(i-1)} \leq z < Z_{(i)}$,

$$R_{n,j}(z) = \frac{j}{T_n(F_n^{-1}F_n(z)) - T_n(F_n^{-1}(F_n(z) - jk_{(n)}^{-1}))} . \qquad (1.4)$$

In the above expression, j is the number of failure intervals used in estimating the failure rate at each fixed time point.

We now give two equivalent expressions for $R_{n,j}$ in terms of R_n , our naive estimator. These two expressions reveal the "smoothing" more transparently. In Section 1.3 we shall show that for $Z_{(i-1)} \leq z < Z_{(i)}$,

$$(R_{n,j}(z))^{-1} = \frac{1}{j} \sum_{m=0}^{j-1} \frac{1}{R_n(Z_{(i-m)})} \qquad (1.5)$$

and that

$$R_{n,j}(z) = \frac{\int_{Z_{(i-1)}}^{Z_{(i)}} R_n(u)dT_n(u)}{\int_{Z_{(i-j)}}^{Z_{(i)}} dT_n(u)} . \qquad (1.6)$$

Since $T_n(t)$ is the total time on test at time t , Equation (1.6) can be regarded as the filtering of $R_n(u)$ through a random window $T_n(u)$, the total time on test at time u. We shall refer to this as the "total time on test random window." In the following section we present considerations which led us to the choice of our smoothing technique. In the sequel, we also introduce some notation which will be used later.

1.2 Justification for Smoothing Technique

In retrospect, we state that a reasonable justification for the choice of our smoothing technique is that it leads us to a tractable distribution-free theory. However, a primary consideration which led us to our smoothing technique was our desire to clarify and to expand the asymptotic results of Singpurwalla (1975). Of particular interest is his

Theorem 4.2, which states that for an exponential distribution, the first difference of the reciprocal of the estimated failure rate converges weakly to a moving average process. He uses this theorem to give a partial justification for an empirical (Box-Jenkins type) time series analysis of the estimated failure rate. Singpurwalla attempts to prove this theorem by using the following theorem (see Section 2):

If F is exponential, then for $t \in \left[\frac{i-1}{k(n)}, \frac{i}{k(n)}\right)$, $i=1,2,\ldots,k(n)$, the "total time on test process" $V_{[k(n)t]} = T_n(Z_{(i-1)})/T_n(Z_{k(n)})$ converges weakly to the Brownian bridge W^o [Billingsley (1968)]. If for any fixed h, $0 < h < 1$, we define the difference operator $\nabla_h : [0,1] \to [h,1]$; $\nabla_h f(x) = f(x) - f(x-h)$, then $\nabla_h^2 W^o$ is a moving average process. Thus, considering the differences of the total time on test process $V_{[k(n)t]}$ becomes quite natural.

We shall see in Section 3 that $\nabla_h V_{[k(n)t]}$, $h = j/n$, is the reciprocal of our randomly smoothed estimator $R_{n,j}(z)$ given by Equation (1.4); this was the rationale behind the choice of our estimator.

We can provide some further insight into the choice of our smoothed estimator if we note that the right-hand side of Equation (1.4) denotes the total number of failures $j \geq 1$ divided by the total time on test between $Z_{(i)}$ and $Z_{(i-j)}$. Thus, our smoothed estimator is merely an extension of our naive estimator, and is obtained by considering a random interval $Z_{(i)} - Z_{(i-j)}$, where $j = [hk(n)]$.

1.3 Equivalence Between the Versions of $R_{n,j}(z)$

It now remains to be shown that $(R_{n,j}(z))^{-1}$ equals the right-hand side of Equation (1.5), and that Equations (1.5) and (1.6) are indeed equivalent.

To see that the former is true, we note [cf. Barlow and Campo (1975)], that

$$\frac{1}{k(n)} (T_n(Z_{(i)}) - T_n(Z_{(i-j)})) = \int_{Z_{(i-j)}}^{Z_{(i)}} \bar{F}_n(u)du$$

and that $R_n(z)$ is specified by Equation (1.1).

In order to verify the equivalence between Equations (1.5) and (1.6), we recall a property of the total time on test. This property [cf. Barlow and Campo (1975)], states that the derivative of the total time on test is the reciprocal of the failure rate.

1.4 Summary

The remainder of this paper is devoted to investigating the distribution theory of $R_{n,h}$, and demonstrating its usefulness.

In Section 2 we present some preliminary results, and in Section 3 we apply these to $R_{n,h}$ when F is exponential. Here we also clarify and expand upon some of the results of Singpurwalla. In Section 4, we consider the distribution theory of $R_{n,h}$ for both finite and large samples, and present a theory for construction confidence bands for r, when r is filtered through the total time on test window.

In Section 5, we discuss the calculation of critical values for use in constructing confidence bands. In Section 6 we illustrate our technique for estimating the failure rate of AC generators, based on some real life data. In Section 7 we generate failure data from a known distribution and use it to estimate the failure rate. We conclude the paper with some discussion of the technique.

2. Preliminary Distribution Theory

We shall assume the conditions of failure and random withdrawals, as stated in Section 1. We shall also adhere to the notation of Section 1.

The text of our paper is based on the following distribution-free result. Lemma 2.1 is a mild extension of that due to Barlow and Proschan (1969).

Lemma 2.1: For any distribution F $(F(0^-)=0)$ with failure rate $r(\cdot)$, the random variables

$$Y_i = \int_{Z_{(i-1)}}^{Z_{(i)}} r(u)N_n(u)du \ , \qquad i=1,2,\ldots,k(n)$$

are independently distributed with density e^{-y}.

Proof: The proof follows that of Barlow and Proschan, by conditioning over the values of L_i and by assuming independence between the L_i and the X_i. //

We recall that $k(n)$ equals the observed number of failures from a collection of n identical items.

Theorem 2.2: When $n \to \infty$,

$$\left\{\frac{1}{\sqrt{k(n)}}\left(\int_0^{Z_{[tk(n)]}} r(u)N_n(u)du - tk(n)\right), \ 0 \le t \le 1\right\} \xrightarrow{\mathcal{D}} \{W(t) \ , \ 0 \le t \le 1\} \ ,$$

where $\xrightarrow{\mathcal{D}}$ indicates convergence of random functions and W is the standard Brownian motion [Billingsley (1968)].

Proof: From Lemma (2.1),

$$\int_0^{Z_{[tk(n)]}} r(u)N_n(u)du = \sum_{i=1}^{[tk(n)]} Y_i \ ; \qquad (2.1)$$

the result follows from Donsker's Theorem (Billingsley, p. 68). //

Corollary 2.3: When F is exponential with $r(u) = \lambda$, $u \ge 0$,

$$\left\{\sqrt{k(n)}\left(\frac{\lambda T_n(Z_{[tk(n)]}}{k(n)} - t\right), \ 0 \le t \le 1\right\} \xrightarrow{\mathcal{D}} \{W(t) \ , \ 0 \le t \le 1\} \ .$$

For completeness, we also present the following well-known theorem [Pyke (1969), Shorack (1972)].

Theorem 2.4: When F is exponential with $r(u) = \lambda$, $u \geq 0$,

$$\left\{\sqrt{k(n)}\left(\frac{T_n(Z_{[tk(n)]})}{T_n(Z_{(k)})} - t\right), \ 0 \leq t \leq 1\right\} \xrightarrow{\mathcal{D}} \{W^o(t) \ , \ 0 \leq t \leq 1\} \ , \qquad (2.2)$$

where W^o is the Brownian bridge [Billingsley (1968)].

Proof: The proof follows from the fact that

$$T_n(Z_{[tk(n)]}) = \frac{1}{\lambda} \sum_{i=1}^{[tk(n)]} Y_i$$

and that

$$\sum_{i=1}^{j} Y_i \Big/ \sum_{i=1}^{k} Y_i \ , \qquad j=1,2,\ldots,k-1$$

are distributed as the $(k-1)$ order statistics from a uniform $[0,1]$ distribution [Karlin (1966)]. These in turn converge to a Brownian bridge.//

Theorem 2.5:

(a) $$\int_0^{Z_{[tn(k)]}} (R_n(u) - r(u))N_n(u)du = \sum_{i=1}^{[tn(k)]} (1 - Y_i)$$

and

(b) $$\left\{\frac{1}{\sqrt{k(n)}} \int_0^{Z_{[tn(k)]}} (R_n(u) - r(u))N_n(u)du \ , \ 0 \leq t \leq 1\right\} \xrightarrow{\mathcal{D}} \{W(t) \ , \ 0 \leq t \leq 1\} \ .$$

Proof: We first note that

$$\int_{Z_{(i-1)}}^{Z_{(i)}} R_n(u)N_n(u)du = \frac{1}{T_n(Z_{(i)}) - T_n(Z_{(i-1)})} \int_{Z_{(i-1)}}^{Z_{(i)}} N_n(u)du = 1 \ .$$

Part (a) of the theorem follows from this and Equation (2.1). Part (b) of the theorem follows from Donsker's Theorem. //

In terms of the "total time on test process," part (b) of the above theorem gives us

$$\left\{\frac{1}{\sqrt{k(n)}}\left(\int_0^{Z_{[tk(n)]}} R_n(u)dT_n(u) - \int_0^{Z_{[tk(n)]}} r(u)dT_n(u)\right),\ 0 \le t \le 1\right\}$$
$$= \left\{\frac{1}{\sqrt{k(n)}}\left(\sum_{i=1}^{[tk(n)]} (1 - Y_i)\right),\ 0 \le t \le 1\right\} \xrightarrow{\mathcal{D}} \{W(t)\ ,\ 0 \le t \le 1\} . \tag{2.3}$$

3. Asymptotic Theory for Exponential Lifetimes

In this section, and the following one, we derive some distribution theory for our smoothed estimator $R_{n,h}$. We do this by taking differences of the expressions derived in Section 2.

The difference operator ∇_h was defined in Section 1.2. For any fixed h , $0 < h < 1$, ∇_h is a continuous function in Skorohod's D-topology, and thus when ∇_h is applied to the statement of Corollary 2.3, the continuous mapping theorem (Billingsley, Theorem 5.1) applies to give

$$\left\{\sqrt{k(n)}\left(\frac{\nabla_h \lambda T_n(Z_{[tk(n)]})}{k(n)} - \nabla_h t\right),\ h \le t \le 1\right\} \xrightarrow{\mathcal{D}} \{\nabla_h W(t)\ ,\ h \le t \le 1\} .$$

Since $\nabla_h t = h$, and since Equation (1.3) implies

$$R_{h,n}(Z_{([tk(n)]-1)}) = \frac{[hk(n)]}{\nabla_h T_n(Z_{[tk(n)]})} ,$$

we have proved

Theorem 3.1: When F is exponential with $r(u) = \lambda$, $u \ge 0$,

$$\left\{\sqrt{k(n)}\left(\frac{1}{R_{n,h}(Z_{([tk(n)]-1)})} - \frac{1}{\lambda}\right),\ h \le t \le 1\right\} \to \{\nabla_h \frac{1}{h} W(t)\ ,\ h \le t \le 1\} ,$$

which is equivalent to

$$\left\{\sqrt{k(n)}\left(\frac{1}{R_{n,h}(Z_{[tk(n)]})} - \frac{1}{\lambda}\right),\ 0 \le t \le 1\right\} \to \{\nabla_h \frac{1}{h} W(t)\ ,\ h \le t \le 1\} .$$

An analogous result can be proved by taking differences of (2.2). Singpurwalla (1975) does this; however, he uses a sequence of difference operators $\nabla_{1/n} \to 0$. In this context, Theorem 5.5 of Billingsley must be used. If we apply the operator $\nabla_{1/n}$ to the statement of Corollary 2.3, and assume that $k=n$, we obtain

$$\left\{\sqrt{n}\left(\frac{\nabla_{1/n}\lambda T_n(Z_{[tn]})}{n} - \nabla_{1/n}t\right), \frac{1}{n} \le t \le 1\right\} \overset{\mathcal{D}}{\to} 0 .$$

Since $\nabla_{1/n}t = 1/n$, and since Equation (1.1) implies

$$R_n(Z_{([tn]-1)}) = \frac{1}{\nabla_{1/n}T_n(Z_{[tn)})} ,$$

we have

$$\left\{\frac{1}{\sqrt{n}}\left(\frac{1}{R_n(Z_{[tn]})} - \frac{1}{\lambda}\right), \frac{1}{n} \le t \le 1\right\} \overset{\mathcal{D}}{\to} 0 .$$

The above result is result is consistent with the fact that the naive estimator R_n^{-1} has a non-degenerate limiting distribution [cf. Sethuraman and Singpurwalla (1977)]. Similarly, the limit in Theorem 4.2 of Singpurwalla should be 0.

3.1 Time Series Analysis of Failure Rates

The value of performing an empirical time series analysis on the estimated failure rate was demonstrated by Singpurwalla (1975). A goal there was to obtain a limiting process of the form: a trend plus a Box-Jenkins type ARIMA process. For complete samples, it follows from Theorem 3.1 that

$$(R_{n,h}(Z_{[tn]}))^{-1} = \lambda^{-1} + (\sqrt{n}\ h)^{-1}\nabla_h W(t) ;$$

that is, a constant plus the increments of a Brownian motion, which is similar to discrete white noise. An analogous result is observed by Sethuraman and Singpurwalla for the naive estimator R_n . Thus it appears that the desired goal of obtaining a limiting process of the form: a trend plus a Box-Jenkins ARIMA process is not possible, unless some additional structure is assumed.

However, we conjecture that the above goal could possibly be realized if the sample size is finite, and if the withdrawals are dependent on lifetimes. We are studying this possibility in the light of the following ramifications:

1. If the lifetimes are exponential, they do not age; this makes it difficult to imagine how withdrawals depend on the lifetime.

2. Without making restrictions on the life distributions, it is impossible to determine from the data whether or not withdrawals are dependent on lifetimes [Miller (1976)].

4. Distribution Theory for General Lifetimes

As in Section 3, we apply the difference operator ∇_h to Equation (2.3), and noting that $h = j(k(n))^{-1}$, we obtain

$$\int_{Z_{([tk(n)]-j)}}^{Z_{([tk(n)]}} R_n(u)dT_n(u) - \int_{Z_{([tk(n)]-j)}}^{Z_{([tk(n)]}} r(u)dT_n(u) = \sum_{i=[tk(n)]-j+1}^{[tk(n)]} (1-Y_i) . \tag{4.1}$$

We next define a "randomly smoothed" version of r,

$$r_{T_{n,j}}(z) = \frac{\int_{Z_{(i-j)}}^{Z_{(i)}} r(u)dT_n(u)}{T_n(Z_{(i)}) - T_n(Z_{(i-j)})} , \quad Z_{(i-j)} \leq z < Z_{(i)} . \tag{4.2}$$

Combining Equation (4.2) with Equation (1.6) we obtain, for $Z_{(i-1)} \leq z < Z_{(i)}$,

$$\left(T_n(Z_{(i)})-T_n(Z_{(i-j)})\right)\left(R_{n,j}(z)-r_{T_{n,j}}(z)\right) = \sum_{m=i-j+1}^{i} (1-Y_m) ,$$

where Y_m, $m=1,2,\ldots,k(n)$ are independent exponential random variables with mean 1.

The analogous result for $R_{n,h}$ is

$$\left(T_n(F_n^{-1}F_n(z)) - T_n(F_n^{-1}(F_n(z)-h))\right)\left(R_{n,h}(z) - r_{T_{n,[hk(n)]}}(z)\right)$$

$$= \sum_{m=[k(n)(F_n(z)-h)]+1}^{[k(n)F_n(z)]} (1-Y_m) = \left. \sum_{m=[k(n)(t-h)]+1}^{[k(n)t]} (1-Y_m) \right|_{t=F_n(z)}$$

$$= \nabla_h \left. \sum_{m=0}^{[k(n)t]} (1-Y_m) \right|_{t=F_n(z)} .$$

By Donsker's Theorem and the continuous mapping theorem we obtain

$$\left\{\frac{1}{\sqrt{k(n)}} \left(T_n(F_n^{-1}(t)) - T_n(F_n^{-1}(t-h))\right)\left(R_{n,h}(F_n^{-1}(t)) - r_{T_{n,[hk(n)]}}(F_n^{-1}(t))\right) ,\right.$$

$$\left. h \le t \le 1\right\} \overset{\mathcal{L}}{\to} \left\{\nabla_h W(t),\ h \le t \le 1\right\} . \tag{4.4}$$

Equation (4.3) can be used to find the finite sample confidence bounds for the "randomly smoothed" failure rate function $r_{T_{n,j}}$. The asymptotic ($n\to\infty$) confidence bounds for $r_{T_{n[hk(n)]}}$ can be obtained from Equation (4.4).

Let $C^+_{j,k,\alpha}$ be the critical value such that

$$P\left\{\sup_{j\le i\le k(n)} \sum_{m=i-j+1}^{i} (Y_m-1) \le C^+_{j,k,\alpha}\right\} = 1-\alpha ,$$

and define $C^-_{j,k,\alpha}$ in a similar manner for the infimum. Then, a $100(1-\alpha)\%$ upper confidence bound for $r_{T_{n,j}}$ is obtained from Equation (4.3) as

$$R_{j,n}(z) + C^+_{j,k,\alpha}\left(T_n(F_n^{-1}(F_n(z)) - T_n(F_n^{-1}(F_n(z)-jk^{-1}))\right)^{-1} \ge r_{T_{n,j}}(z) ,$$

for all z, $Z_{(j-1)} \le z < Z_{(k)}$.

The above simplifies to

$$R_{j,n}(z)(1 + C^{+}_{j,k,\alpha}/j) \ , \tag{4.5}$$

for $Z_{(j-1)} \leq z < Z_{(k)}$. A lower confidence bound is similarly defined.

An approximate $100(1-\alpha)\%$ confidence bound is obtained from Equation (4.4) as

$$R_{h,n}(z) + \sqrt{k(n)}\, C^{+}_{h,\infty,\alpha} \left(T_n(F_n^{-1}(F_n(z))) - T_n(F_n^{-1}(F_n(z)-h))\right)^{-1}$$

$$\geq r_{T_{n,[hk(n)]}}(z) \ ,$$

for all z , $Z_{([hk(n)]-1)} \leq z < Z_{(k)}$, where $C^{+}_{h,\infty,\alpha}$ is the critical value such that

$$P\left\{\sup_{h\leq t\leq 1} \nabla_h W(t) \leq C^{+}_{h,\infty,\alpha}\right\} = 1-\alpha \ .$$

A lower confidence bound and a two-sided confidence bound can be similarly obtained.

5. Critical Values for Confidence Bounds for Randomly Smoothed Failure Rate

To the best of our knowledge, there do not presently exist analytical methods for calculating $C^{+}_{j,k,\alpha}$ or $C^{+}_{h,\infty,\alpha}$. We therefore used a Monte Carlo method to calculate selected values needed in Sections 6 and 7; these are:

$$C^{-}_{1,55,.10} = -.9981 \qquad C^{+}_{1,55,.10} = 5.279$$

$$C^{-}_{1,55,.05} = -.9990 \qquad C^{+}_{1,55,.05} = 6.008$$

$$C^{-}_{5,55,.10} = -4.06 \qquad C^{+}_{5,55,.10} = 7.887$$

$$C^{-}_{5,55,.05} = -4.21 \qquad C^{+}_{5,55,.05} = 9.02 \ .$$

The values are based on one run consisting of 20,000 replicates. On a second run, also consisting of 20,000 replicates, it was determined that the four two-sided confidence bands using the above values would have confidence values approximately equal to .8035, .8988, .8094, and .9052, respectively. Thus, the one-sided critical values can be used for approximate two-sided confidence bands.

6. Application to Failure Data on AC Generators

We illustrate our estimation and smoothing technique by considering some failure data on AC generators reported in NAILSC Report ILS 04-21-72. This data was previously considered by Castellino and Singpurwalla (1973), and by Singpurwalla (1975). This data is presented in Table 1 and consists of failure and removal (withdrawal) times.

In Figure 1 we present the estimated failure rate using Equation (1.1). The confidence bounds using Equation (4.5) are so wide that they are not useful. For instance, a 90% lower bound is $(1 - .9981)R(z) = .0019R(z)$, whereas a 90% upper bound is $(1 + 5.279)R(z) = 6.279R(z)$.

In Figure 2 we show the estimated failure rate using Equation (1.4), $R_{5,55}$. We also show the 90% upper and lower confidence bounds for $r_{T_{n,5}}(\cdot)$, the randomly smoothed version of the true failure rate. We remark here that the estimator $R_{5,55}$ was calculated using the current interval plus the two leading and the two lagging intervals. Clearly, the smoothing leads us to narrower bands than those obtained by considering single intervals. The main issue here is that the confidence bands pertain to a version of r, rather than to r itself. Finally, the periodic behavior of the failure rate, observed by Singpurwalla (1975), is also exhibited by $R_{5,55}$.

7. Application to Monte Carlo Example and a Discussion

In order to assess the performance of the above estimation procedure in a known situation, we consider failure data generated from the failure rate function

$$r(t) = \left(1 + \frac{t}{8} + \sin\frac{\pi t}{4}\right) \Big/ 12 \ . \tag{7.1}$$

In Table 2 we present the failure times, and in Figure 3 we show a plot of $r(t)$. In Figure 4 we present a plot of the estimated failure rate $R_{5,55}$, together with the 90% upper and lower confidence limits for $r_{T_{n,5}}$. In Figures 5 and 6, we show the "randomly smoothed" failure rate functions $r_{T_{n,1}}$ and $r_{T_{n,5}}$, respectively. If we superimpose Figures 4 and 6, we see that $r_{T_{n,5}}$ falls well within the confidence band.

7.1 Concluding Remarks

We would like to give some interpretation to the randomly smoothed failure rate r , smoothed by the total time on test window.

Suppose that we first consider R_n our naive estimator, which gives a constant value over failure intervals. Thus, it is logical to consider this as an estimator of r averaged over the interval of interest; however, this is essentially $r_{T_{n,1}}$. Unless some regularity conditions are imposed on r , it will be impossible to obtain confidence bounds for r over the interval. Thus, in order to obtain confidence bounds, we consider $r_{T_{n,1}}$ instead of r . However, the confidence bound on $r_{T_{n,1}}$ is so wide that we smooth over more than one interval. Thus in essence, it is a small step to go from smoothing over one interval to smoothing over several intervals. Finally, even though the smoothing window is random (it depends on the data), it can be completely specified by the failure data and is thus completely known for the purpose of interpreting the confidence band.

TABLE 1

FAILURE DATA FOR AC GENERATORS

Failure Times

1.0	1.3	3.0	3.0	3.8	5.2	10.6
24.6	29.5	30.6	34.5	38.9	40.9	43.2
70.0	71.6	100.5	141.1	160.8	164.0	167.0
200.3	206.2	212.5	229.0	252.8	252.8	272.7
274.3	275.3	282.2	403.8	435.3	442.0	444.0
466.6	474.6	495.0	500.0	502.5	509.0	510.8
520.7	523.0	599.0	666.6	676.0	677.0	703.6
744.8	827.1	852.0	861.0	950.4	1097.3	

Censoring Times (Withdrawal Times)

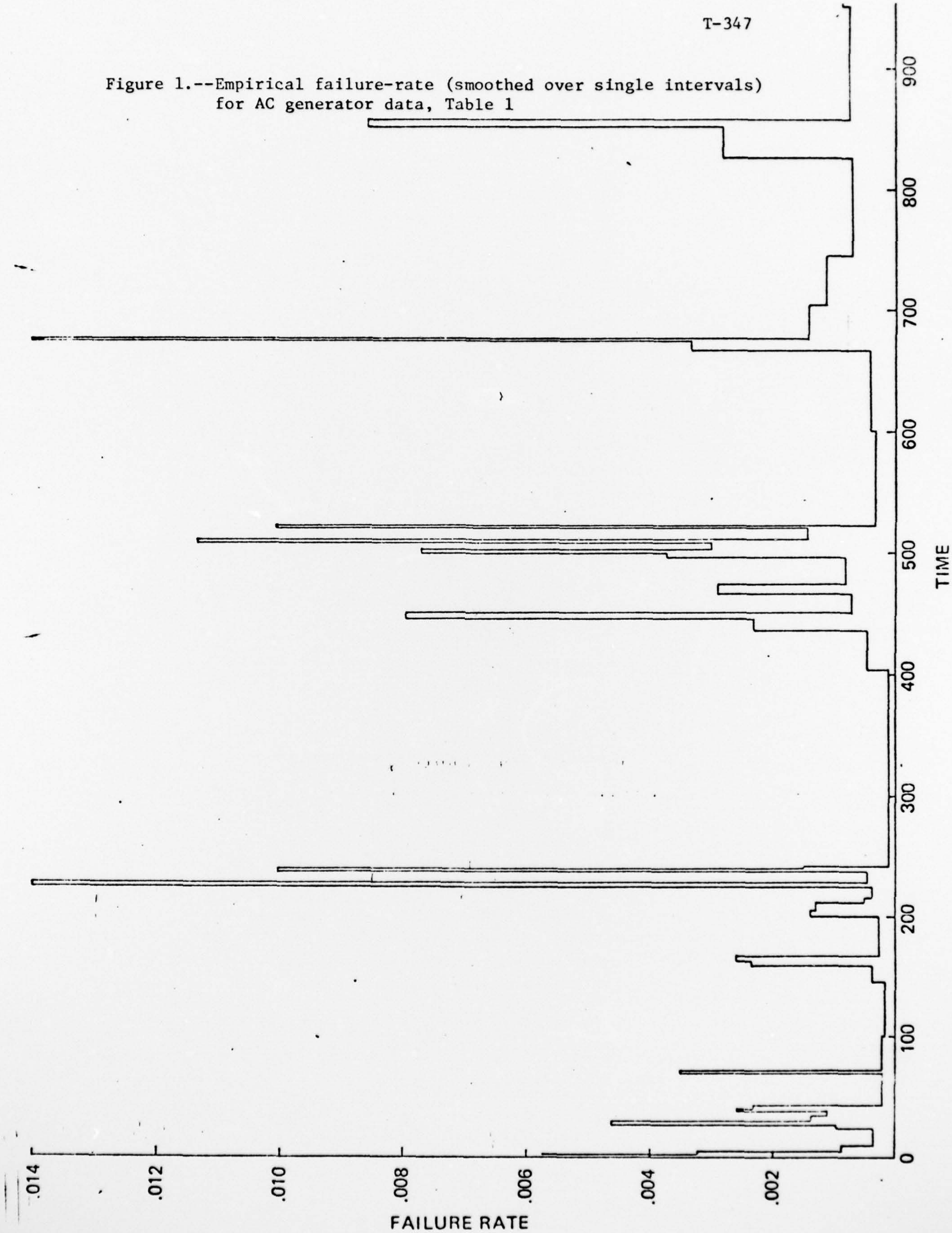
0.5	0.5	1.0	1.3	1.5	2.6	5.2
6.1	6.2	6.4	6.5	12.0	12.7	16.2
17.0	17.2	19.2	19.5	20.1	23.7	24.7
26.4	28.6	34.5	36.6	38.9	43.4	45.6
50.6	53.7	58.4	59.3	62.6	64.9	65.9
67.3	69.3	72.6	72.6	74.7	74.8	79.2
79.6	80.0	81.0	83.4	83.4	84.3	86.9
89.0	90.1	90.5	91.3	92.8	100.2	105.4
107.1	109.0	111.8	112.1	113.8	117.0	125.9
126.6	131.2	131.5	132.0	134.3	135.0	140.3
142.1	149.2	149.7	157.1	158.4	161.0	164.0
164.0	166.4	173.5	174.0	184.4	187.2	191.6
211.3	216.5	218.2	225.3	228.9	233.3	237.4
243.1	261.0	265.3	265.7	268.0	268.6	270.5
272.7	274.6	274.9	275.2	283.5	292.0	301.8
306.2	314.0	319.0	320.7	321.4	328.7	330.6
338.2	347.8	349.0	355.1	359.8	371.8	373.6
381.0	382.8	389.0	393.2	402.6	403.7	406.1
406.8	411.3	414.2	428.8	435.3	441.0	446.1
461.1	465.7	466.6	479.6	498.8	505.8	510.8
512.0	514.8	518.3	585.2	605.2	606.4	617.4
627.7	627.7	630.0	645.9	674.4	684.3	685.2
689.9	701.7	704.4	724.4	732.7	752.5	771.2
784.8	784.9	824.4	864.3	881.7	898.7	952.0
957.0	983.1	1064.1	1147.3	1173.6	1762.2	

TABLE 2

MONTE CARLO FAILURE TIMES

.127	.406	.425	.635	.669
1.041	1.132	1.156	1.488	1.788
1.907	1.991	2.056	2.122	2.362
2.366	2.473	2.604	2.900	2.970
3.083	3.372	3.557	3.943	4.012
4.352	5.103	5.615	6.843	6.928
7.280	7.942	8.310	8.311	8.546
8.792	8.930	9.137	9.241	9.561
9.904	9.988	10.375	11.507	11.735
12.748	13.301	13.508	14.435	15.518
16.857	17.266	17.426	17.975	18.259

Figure 1.--Empirical failure-rate (smoothed over single intervals) for AC generator data, Table 1



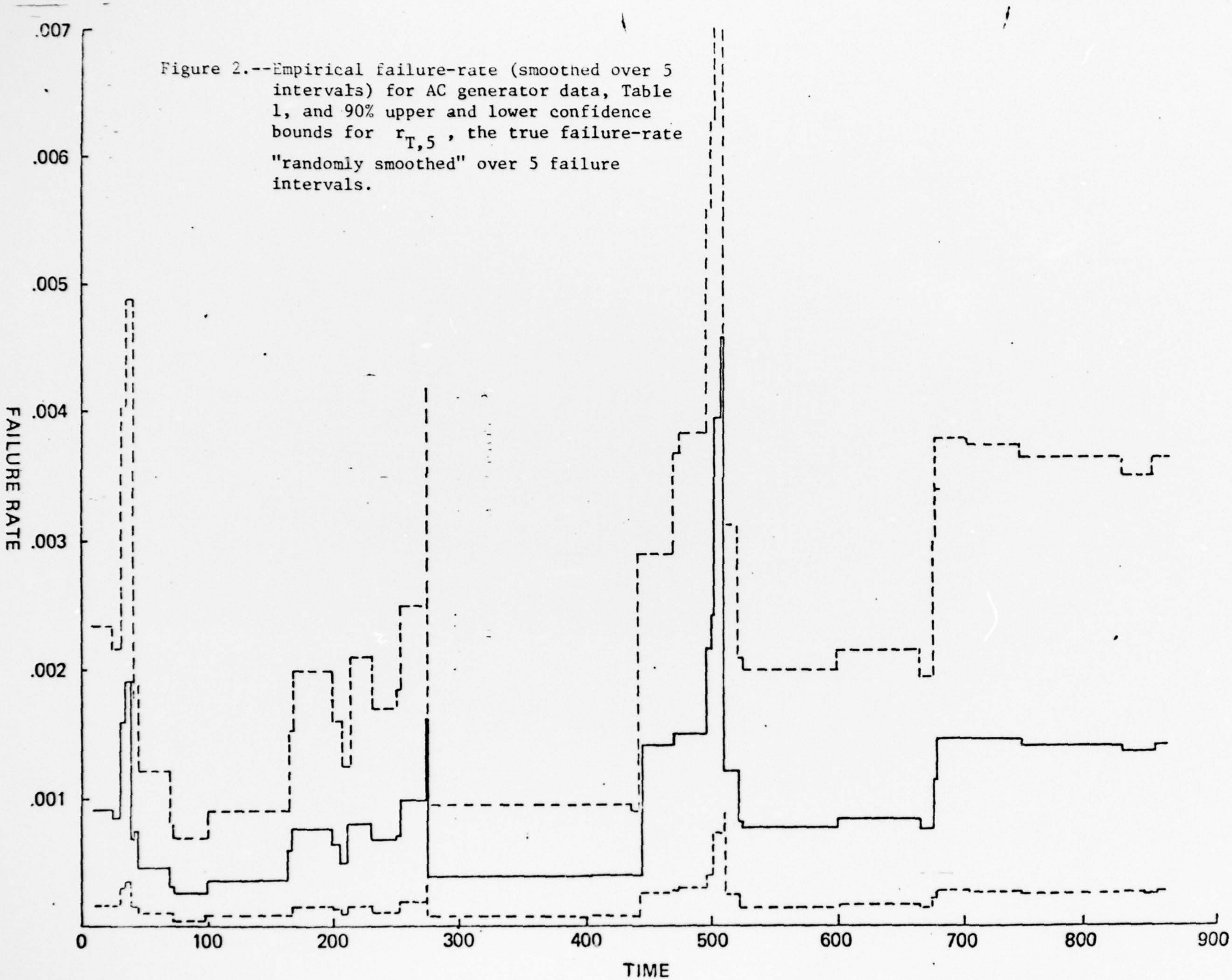


Figure 2.--Empirical failure-rate (smoothed over 5 intervals) for AC generator data, Table 1, and 90% upper and lower confidence bounds for $r_{T,5}$, the true failure-rate "randomly smoothed" over 5 failure intervals.

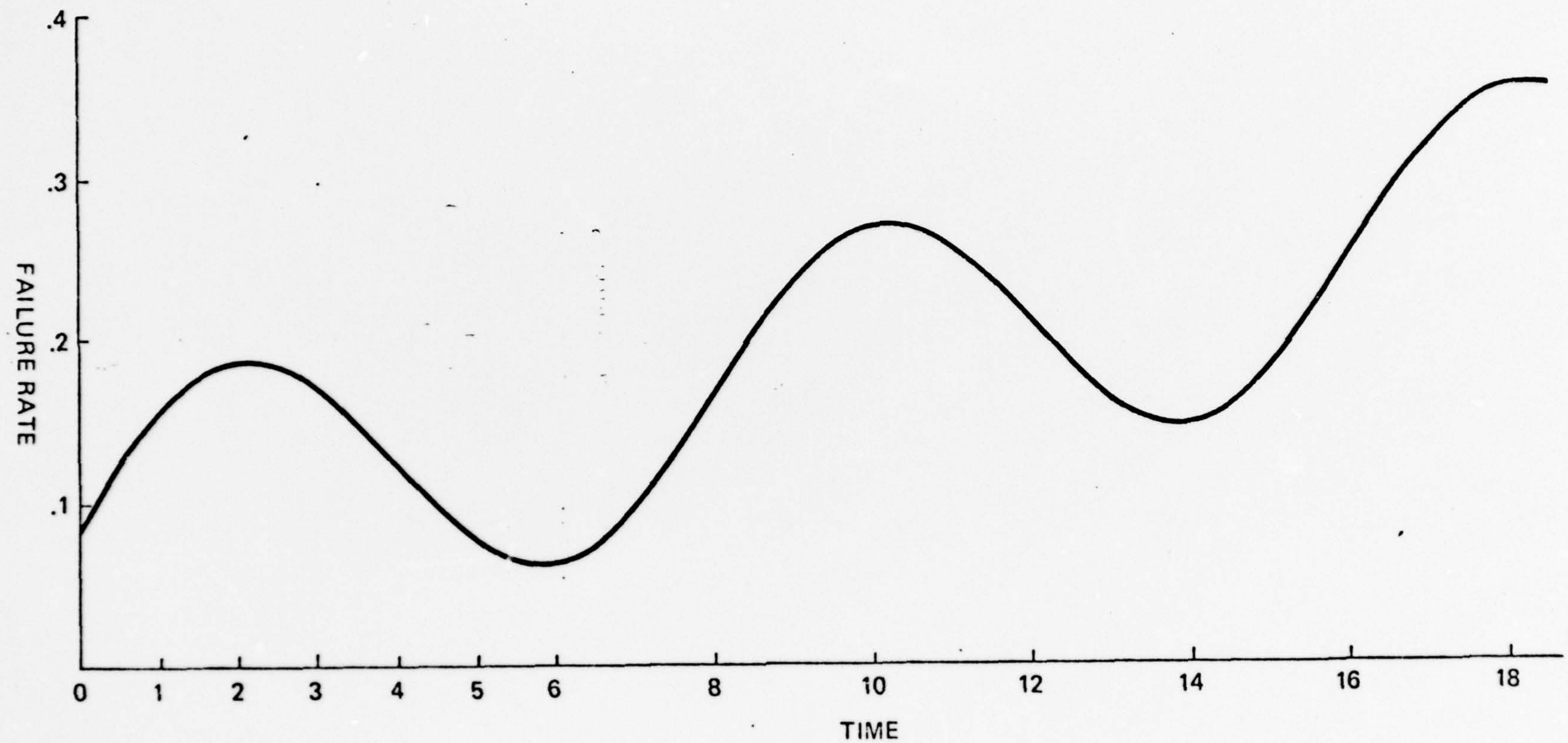
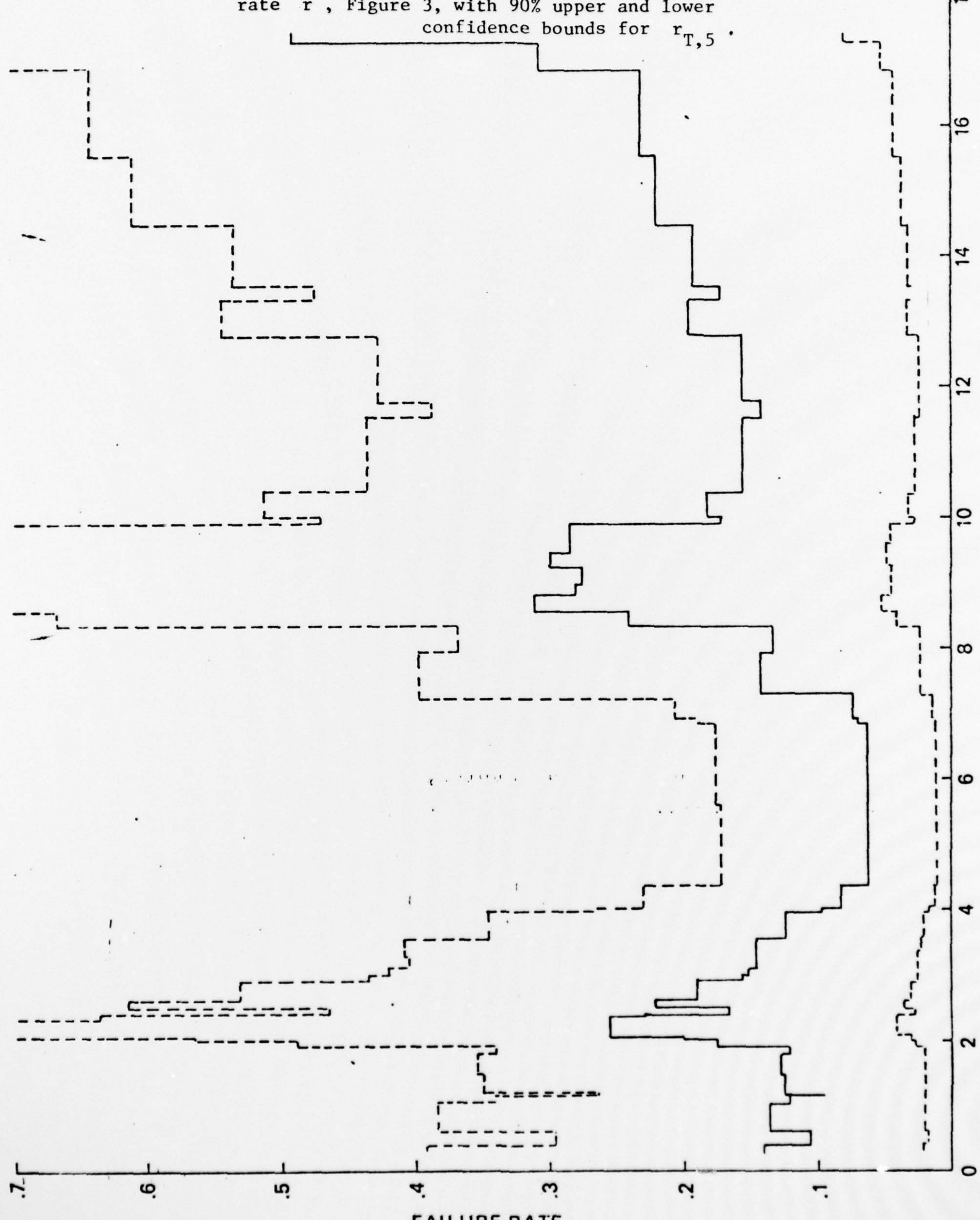


Figure 3.--Failure-rate function, $r(t) = \left(1 + \frac{t}{8} + \sin\frac{\pi t}{4}\right) \Big/ 12$.

Figure 4.--Empirical failure-rate (smoothed over 5 intervals) for data in Table 2 with failure rate r , Figure 3, with 90% upper and lower confidence bounds for $r_{T,5}$.



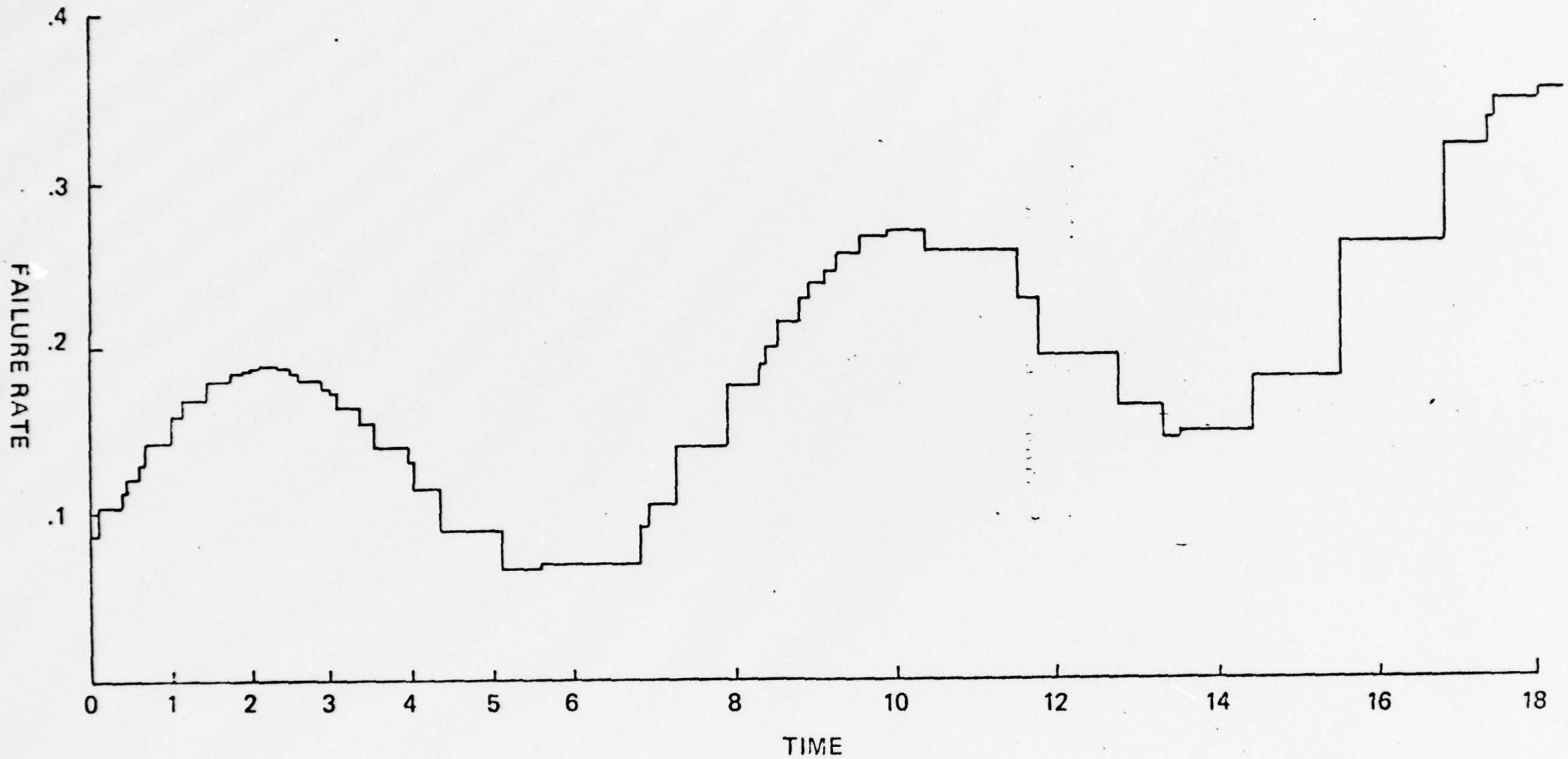


Figure 5.--"Randomly smoothed" failure-rate r , Figure 3; smoothed over single failure intervals using data in Table 2.

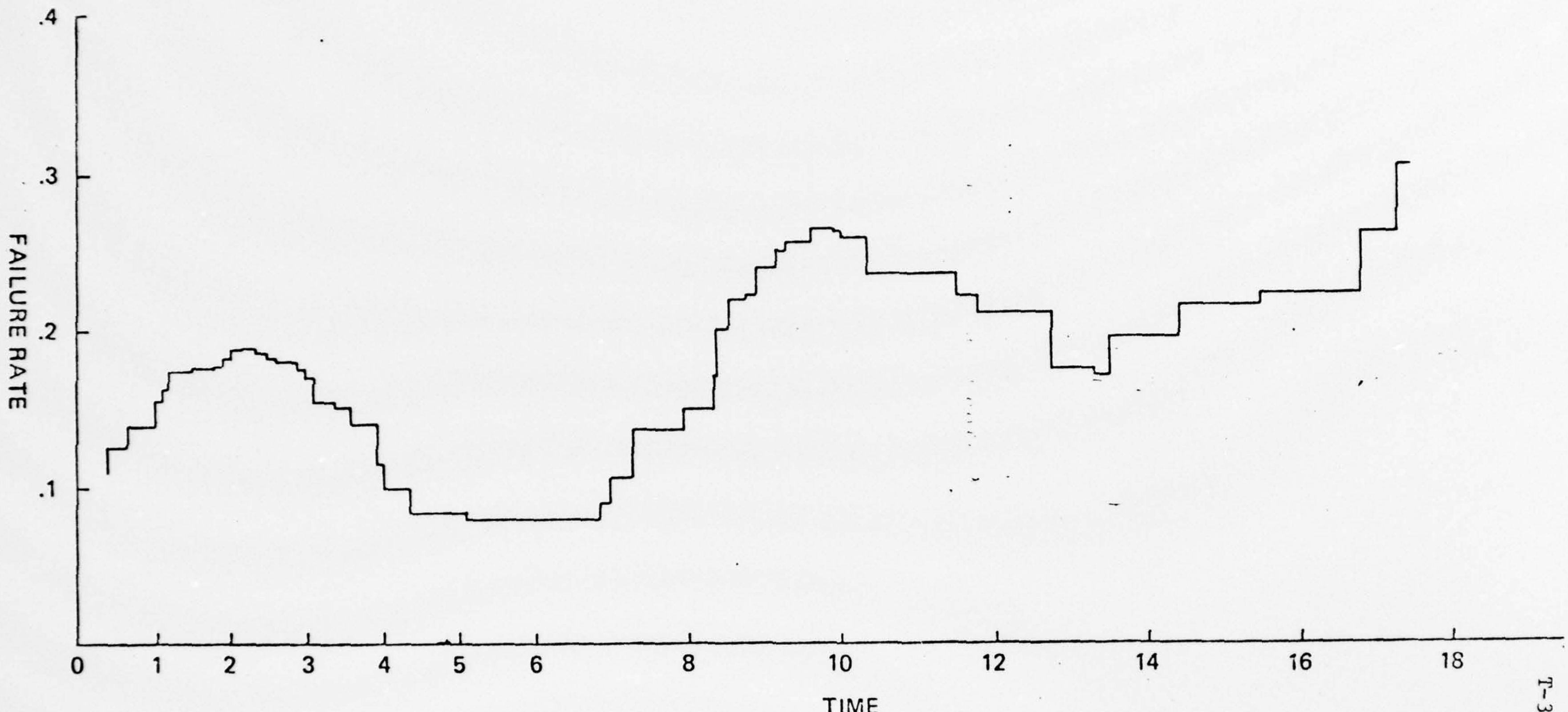


Figure 6.--"Randomly smoothed" failure-rate r , Figure 3; smoothed over 5 failure intervals using data in Table 2.

REFERENCES

BARLOW, R. E. and R. CAMPO (1975). Total time on test processes and applications to failure data analysis. Reliability and Fault Tree Analysis (R. E. Barlow, J. B. Fussell and N. D. Singpurwalla, eds.). SIAM, Philadelphia.

BARLOW, R. E. and F. PROSCHAN (1969). A note on a test for monotone failure rate based on incomplete data. Ann. Math. Statist. **40** 595-600.

BARLOW, R. E. and F. PROSCHAN (1975). Statistical Theory of Reliability and Life Testing. Holt, Rinehart, and Winston, New York.

BILLINGSLEY, P. (1968). Convergence of Probability Measures. Wiley, New York.

CASTELLINO, V. F. and N. D. SINGPURWALLA (1973). On the forecasting of failure rates -- a detailed analysis. Technical Memorandum Serial TM-62197, Program in Logistics, The George Washington University.

KARLIN, S. (1966). A First Course in Stochastic Processes. Academic Press, New York.

MILLER, D. R. (1976). A note on independence of multivariate lifetimes in competing risks models. Ann. Statist., to appear.

NAILSC Report ILS 04-21-72 (1972). Technique for scheduled removal component selection and interval determination. Naval Aviation Integrated Logistics. Support Center.

PYKE, R. (1969). Applications of almost surely convergent constructions of weakly convergent processes. Lecture Notes in Math. **89** 187-200.

SETHURAMAN, J. and N. D. SINGPURWALLA (1977). On a naive estimator of the failure rate. To appear.

SHORACK, G. R. (1972). Convergence of quantile and spacings processes with applications. Ann. Math. Statist. **43** 1400-1411.

SINGPURWALLA N. D. (1975). Time series analysis and forecasting of failure-rate processes. Reliability and Fault Tree Analysis (R. E. Barlow, J. B. Fussell, and N. D. Singpurwalla, eds.). SIAM, Philadelphia.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: UMC-67 2. GOVT ACCESSION NO.: TR-67	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Failure Rate Estimation Using Random Smoothing	5. TYPE OF REPORT & PERIOD COVERED: Interim Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Douglas R. Miller, Nozer D. Singpurwalla	8. CONTRACT OR GRANT NUMBER(s): N00014-76-C-0843
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Statistics Department, University of Missouri, Columbia, Missouri 65201	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 042-358
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Department of the Navy, Arlington, VA 22217	12. REPORT DATE: April 1977
	13. NUMBER OF PAGES: 25
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

402 660

18. SUPPLEMENTARY NOTES

also supported by Contract AT(49-24)-0211, Nuclear Regulatory Commission, and Contract N0014-75-C-0729, Project NR 347-020, Office of Naval Research

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

failure rate function, time-on-test process, random smoothing, time-on-test window, confidence regions, weak convergence, Brownian motion

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)